Acquisition of 500 English Words through a TActile Phonemic Sleeve (TAPS)

Hong Z. Tan, *Fellow, IEEE*, Charlotte M. Reed, Yang Jiao, Zachary D. Perez, E. Courtenay Wilson, Jaehong Jung, Juan S. Martinez, and Frederico M. Severgnini

***Abstract*—Recently, a phonemic-based tactile speech communication system was developed with the goal to transmit speech through the skin for people with hearing impairments and those whose auditory and visual channels are overloaded or compromised. The display, called the TActile Phonemic Sleeve (TAPS), consisted of a 4-by-6 tactor array worn on the dorsal and volar surfaces of the forearm. Earlier work showed that people were able to learn the haptic symbols for 39 English phonemes and reach a mean phoneme recognition rate of 86% correct within one to four hours of training. The current research evaluated the acquisition of up to 500 words using TAPS. A total of 51 participants were trained and tested in three studies with increasing number of phonemes and vocabulary sizes. Individual achievements varied, but the results clearly demonstrate the potential of transmitting any English word using TAPS within a reasonable period of learning. Future work will include increasing the speech transmission rate with TAPS by improving the phonemic codes and reducing the inter-phoneme intervals, addressing the reception of words and sentences composed of strings of tactile phonemes, and assessing the performance of TAPS as a speech communication system for people with severe hearing impairments.**

***Index Terms*—Tactile speech communication, phonemic coding, tactile phoneme identification, tactile word identification, language acquisition, learning rate, haptic symbols for English phonemes.**

I. Introduction

We have always known that speech reception through the sense of touch is possible as demonstrated by natural tactile speech communication methods, yet decades of research and development on sensory substitution have not produced tactile devices with performance levels that match those achieved by natural methods. We tackle this challenge with our ongoing efforts at enabling speech reception through the skin. This article summarizes results from three longitudinal studies to demonstrate tactile word acquisition mediated with a TActile Phonemic Sleeve (TAPS). Our data from fifty-one participants show that it is possible to acquire up to 500 English words on the forearm at a learning rate of roughly 1 word per minute.

Evidence of tactile speech reception can be found in a wealth of literature on natural tactile communication methods [1], [2], particularly the methods used by individuals who are both deaf and blind [3]–[10]. One noteworthy example is the Tadoma method where a user places the hand on a speaker's face, with the thumb over the lips, the fingers spread across the cheek and the little finger on the neck. In the absence of any visual or auditory cues, the Tadoma user obtains tactual information associated with articulatory processes such as mouth opening and air flow via the thumb, muscle tension via the fingers on the cheek, and laryngeal vibration via the little finger. Performance with the Tadoma method has been well documented: After years of learning, Tadoma users can achieve a performance level of roughly 55% correct at receiving consonants and vowels, 40% correct at receiving isolated monosyllabic words, and 80% correct at receiving key words in conversational sentences produced at slow-to-normal speaking rates [8]. Two-way communication rates are estimated to be in the range of 60-80 words per minute (wpm) [11], which is comparable to the rates produced for slow conversational speech [12]. Unlike other tactile speech communication methods that are often used with lip reading, information available to Tadoma users is purely touch-based, thus providing proof of an existing lower bound that can be achieved with a natural (i.e., not device-mediated), touch-based speech communication method. Therefore, the performance levels achieved by experienced Tadoma users can serve as benchmarks against which other natural and device-mediated tactile speech communication methods, including our own, can be compared.

In general, performance levels with devices designed for tactile speech communication do not approach the levels demonstrated by Tadoma users [13]. Prosthetic devices for people with profound deafness have been developed that include tactile aids and cochlear implants. Due to the success of cochlear implants where a significant fraction of implanted adults are able to achieve high levels of speech reception through the implant alone, and some children implanted at a very young age may develop speech and language skills at levels comparable to their normal-hearing peers [14], [15], the number of persons with cochlear implants far outnumbers those using tactile aids.

Manuscript received August 15, 2019; revised January 9, 2020 and February 6, 2020; accepted February 7, 2020. Date of publication February 11, 2020; date of current version December 21, 2020. This work was supported by a Grant from Facebook, Inc. This article was recommended for publication by Associate Editor A. Kapper and Editor-in-Chief L. Jones upon evaluation of the reviewers' comments. *(Corresponding author: Hong Z. Tan.)*

Hong Z. Tan, Yang Jiao, Jaehong Jung, Juan S. Martinez, and Frederico M. Severgnini are with the Haptic Interface Research Laboratory, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 (e-mail: hongtan@purdue.edu; jiao12@purdue.edu; jung137@purdue.edu; mart1304@purdue.edu; fredericomqs@gmail.com).

Charlotte M. Reed, Zachary D. Perez, and E. Courtenay Wilson are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 (e-mail: cmreed@mit.edu; zperez@mit.edu; ecwilson@mit.edu).

This article has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TOH.2020.2973135

1939-1412 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

TABLE I
OVERVIEW OF THE THREE STUDIES

Study	# Participants	# Phonemes	# Words	Prior Publication
I	10 (Naïve)	10	51	[33] (P01-P04 data)
II	24 (Naïve)	39	100	[35]
III	11 (Experienced) + 10 (Naïve)	39	500	(none)

Nonetheless, over the years there have been systematic efforts on the development and testing of wearable tactile aids [16]–[18], including the more successfully commercialized Tactaid devices [19] from Audiological Engineering Corp. (AEC, Somerville, MA). The devices are useful to those for whom cochlear implantation is not an option, and provide a non-surgical alternative to people with hearing impairments. When used alone, the Tactaid devices were capable of conveying useful information regarding environmental sounds [20], but could not be used for understanding speech without the use of lipreading [21]. When used in conjunction with lipreading, the devices provided a limited improvement to sentence reception accuracy with a typical increase of around 10% (the so-called ceiling effect) [20]. The Tactaid devices are no longer available commercially yet the need still exists for persons with profound hearing loss as an option for aiding lipreading and receiving environmental sounds. In addition, tactile communication devices can potentially be useful to individuals who are situationally deaf due to activities such as firefighting, diving, military silent operation and in applications such as virtual and augmented reality.

Recently, work has been undertaken by the authors with the goal of developing a tactile speech communication system which permits recognition of any English word after a short learning period. Several decisions were made early in this work regarding a strategy that would optimize the chances of success of this ambitious goal. Firstly, we decided against the use of Braille symbols as the basic unit. Braille codes can be displayed on the skin using a 6-tactor display. Recognizing the Braille patterns on the forearm requires the user to localize and enumerate all vibrating tactors that correspond to the "on" dots in a Braille code. This would require a sparsely-spaced six tactor layout on the forearm as the two-point discrimination threshold on the forearm is at least 30 mm [22] and our ability to localize tactors on the forearm is limited [23], [24]. An additional challenge is that numerosity judgment on the skin is poor especially with tactors placed on a regular grid [25], [26]. In terms of familiarity with Braille codes, fewer than 10% of the 1.3 million people who are legally blind in the US are Braille readers [27], and the learning of Braille may prove to be time-consuming and challenging for users who experience degenerative visual loss and those of older age [10]. Therefore, a Braille-based system will benefit only a small fraction of the blind population and impose additional learning time on other potential users.

Secondly, we decided against the use of Morse Code as the basic unit. It is time-consuming to become efficient at receiving words with the Morse Code. The code is also inherently slow due to the need to maintain the 1:3 dot-dash duration ratio as well as the space between dots and dashes. Previous research has shown that the ability of two experienced Morse Code operators to receive the code through short (for dot) and long (for dash) vibration patterns is limited to about 20 wpm [28], far below the 60-80 wpm demonstrated by Tadoma users.

Thirdly, we decided against the spectral-based encoding approach based on the cochlear model that has been widely used by most tactile aids for speech communication. As an example, Tactaid VII made by AEC consisted of seven resonant-type vibrators. The acoustic signal of speech was processed through an array of bandpass filters with increasing center frequencies. The outputs of these filters were rectified and used to modulate the amplitudes of the corresponding vibrators [29]. One problem with the Tactaid VII is that the tactors all vibrated at the same fixed resonant frequency, and such a "homogeneous" display is prone to masking (a perceptual phenomenon that refers to a reduced sensitivity to one signal in the presence of stronger signals nearby). There is also the additional challenge of token variations both within and across speakers, posing an extra burden on the user to learn to categorize tactile sensations despite the sometimes large variations.

Our strategy is to encode English phonemes into perceptually distinct haptic symbols that can be combined to "sound out" a word. We reasoned that with the recent advancement of automatic speech recognition technologies and text-to-speech transcription, it is possible to convert any spoken or written English into phoneme streams that can be displayed on the forearm. For faster speech transmission, we decided to use phonemes as the basic units instead of letters of the alphabet because the number of phonemes in any English word is always less than or equal to the number of letters. The tradeoff here is that the user needs to learn 39 haptic symbols associated with the 39 English phonemes, as opposed to only 26 symbols for letters. Our earlier work demonstrated, however, that people can indeed learn the 39 haptic symbols for phonemes with a mean recognition rate of 86% correct within one to four hours of training [30].

There are two long-term objectives of our research program. First, we are interested in the information transmission rates achievable with our phonemic-based tactile speech communication system. By leveraging the English language knowledge and skills of our participants, we hope to shorten the time required for the reception of continuous speech materials. Second, we envision our phonemic-based tactile speech communication system to be used by people with all levels of sensorimotor capabilities, including those who are deaf or deaf-and-blind. We are therefore focusing on testing participants with normal vision and hearing before addressing the specific language skills of people with hearing and/or visual impairments. The present report is an extension of the work presented in Reed *et al.* [30] where we presented the tactile coding scheme and results on *phoneme* identification. We report our findings on *word* recognition using the same haptic

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

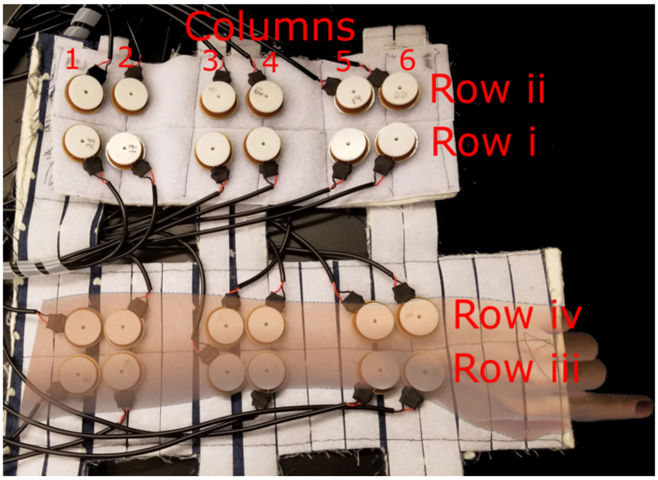

Fig. 1. Tactor layout of TAPS. The row and column numbering is the same as that used in Fig. 1 and Table I of [30] for ease of reference. The superimposed hand and arm image indicates the placement of rows iii and iv under the forearm (volar side), and rows i and ii above the forearm (dorsal side).

symbols representing English phonemes. In a series of three studies (see Table I for an overview), the vocabulary size was increased from 51 to 100 to 500 words, and the efficacy of phoneme-based (bottom-up) and word-based (top-down) learning approaches was compared. A total of 51 participants were recruited across studies (with some taking part in multiple studies). The results provide strong evidence that tactile speech communication is achievable within a reasonable amount of learning time.

II. General Methods

This section presents the general methods that are common to the three studies reported in this article. Information that is unique for each study is included in subsequent sections for the individual studies.

A. TAPS

The haptic interface is called **TA**ctile **P**honemic **S**leeve, or TAPS. It consists of a 4-by-6 tactor array worn on the left forearm (Figure 1). There are six tactors in the longitudinal direction (elbow to wrist) and four tactors in the transversal direction (ring around the forearm). As seen in Figure 1, the 24 tactors are arranged in six groups of four, with three clusters on both the dorsal and volar sides of the forearm. To "wear" the interface, the participant places the left forearm on the lower half of the tactor array (see Figure 1) with the volar side facing down, wraps the upper half of the tactor array on top of the dorsal forearm, and fastens the gauntlet with Velcro straps.

The actuator is a wide-bandwidth tactor (Tectonic Elements, Model TEAX13C02-8/RH, Part #297-214, sourced from Parts Express International, Inc.). A MOTU 24Ao audio device (MOTU, Cambridge, MA, USA) was used for delivering 24 channels of audio waveforms to the 24 tactors through custom-built stereo audio amplifiers. A Matlab program running on a desktop computer generated the multi-channel waveforms and ran the experiments. With this setup, the tactors can be driven independently with programmable waveforms.

Fig. 2. Experimental setup.

During all experiments, the participant sat comfortably in front of a computer monitor and wore noise-reduction earphones to block any sounds from the tactors (see Figure 2). The elbow-to-wrist direction of the left arm was adjusted to be roughly parallel to the torso. The participant used the right hand to operate the computer keyboard and mouse. Several gauntlets were constructed using different fabric materials during the experiments reported here, but the layout of the tactor array on the forearm was kept the same.

B. Haptic Symbols for Phonemes and Words

Thirty-nine haptic symbols were developed for the 39 English phonemes, i.e., distinct sounds of spoken English [31]. Table II shows the 24 consonants and 15 vowels making up the phonemes in the form of our own capital letter symbols, the corresponding International Phonetic Alphabet (IPA) symbols, and sample words containing the corresponding phonemes. We used capital letter symbols for ease of programming and data storage in the text format. As an example, the phoneme transcription for "ace" was "AY"–"S."

Each symbol consists of vibrotactile patterns using one or more of the 24 tactors. The mapping of the phonemes and haptic symbols incorporates the articulatory features of the sounds of the English language, balanced by the need to maintain the distinctiveness of the 39 haptic symbols. The stimulus properties included *amplitude* (in dB sensation level, or dB above individually-measured detection thresholds), *frequency* (single or multiple sinusoidal components), *waveform* (sinusoids with or without modulation), *duration* (100 and 480 ms for short and long, respectively), *location* (place of activation along the TAPS array), *numerosity* (single tactor activation or multiple tactors turned on simultaneously or sequentially), and *movement* (smooth apparent motion or discrete saltatory motion varying in direction, spatial extent, and/or trajectory). Examples of the use of articulatory features to construct the phonemes include the use of location on the array to map place of articulation (e.g., front sounds are presented near the wrist and back sounds near the elbow), the use of unmodulated versus modulated waveforms to distinguish voiceless and voiced cognate pairs (i.e., vibrotactile modulation was used to encode vocal-fold vibration), and the use of short and long signal

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

TABLE II
THE THIRTY-NINE ENGLISH PHONEMES USED IN THE PRESENT STUDY

Our Symbol	IPA	Example Word	Our Symbol	IPA	Example Word	Our Symbol	IPA	Example Word	Our Symbol	IPA	Example Word
EE	/i/	**mee**t	UU	/U/	h**oo**d	K	/k/	**k**ey	ZH	/ʒ/	a**z**ure
AY	/ei/	m**a**te	UH	/Λ/	h**u**t	P	/p/	**p**ay	CH	/tʃ/	**ch**ew
OO	/u/	m**oo**d	OE	/OU/	b**oa**t	T	/t/	**t**ea	J	/dʒ/	**j**eep
I	/aI/	m**igh**t	OY	/ƆI/	b**oy**	B	/b/	**b**ee	H	/h/	**h**e
AE	/ae/	m**a**t	OW	/aU/	p**ou**ch	G	/g/	**g**uy	R	/r/	**r**ay
AH	/ɑ/	f**a**ther	D	/d/	**d**o	F	/f/	**f**ee	L	/l/	**l**ie
AW	/Ɔ/	b**ough**t	M	/m/	**m**e	SH	/ʃ/	**sh**e	Y	/j/	**y**ou
EH	/Ɛ/	m**e**t	S	/s/	**s**ee	TH	/Θ/	**th**ink	N	/n/	**n**ew
ER	/ɚ/	b**ir**d	W	/w/	**w**e	V	/v/	**v**oice	NG	/ŋ/	si**ng**
IH	/I/	b**i**d	DH	/ð/	**th**e	Z	/z/	**z**oo			

TABLE III
THE FIFTY-ONE ENGLISH WORDS USED IN STUDY I (C = CONSONANT; V = VOWEL)

Group 1 (24 words)				Group 2 (27 words)			
CV/VC	CV/VC	CVC	CVC	CV/VC	CV/VC	CVC	CVC
ace	say	dame	maim	coup	they	cake	seek
aid	see	deed	mead	die	thy	came	side
day	sue	deem	mood	kay	way	dime	wade
do		doom	moose	key	we	make	wake
may		dude	same	my	why	meek	weed
me		mace	seam	sigh	woo	mime	wide
moo		made	seed	thee		sake	womb

durations for distinguishing brief plosive bursts from longer fricative noises, respectively.

To further differentiate consonants and vowels, all haptic symbols for consonants occur at distinct locations on the forearm, and those for vowels involve simulated movement sensations (e.g., from the wrist to the elbow for the "OO" sound). Details of the phoneme mapping strategies and the resultant haptic symbols can be found in [30], with supplemental materials detailing the more complex waveforms for vowels.

To display an English word, the haptic symbols corresponding to the phonemes making up the word were delivered in sequence, with a temporal gap (inter-phoneme interval) inserted between phonemes. The word duration varied from roughly 1 to 2 s.

C. Calibration of Perceived Intensity

In order for the haptic symbols to be well perceived despite individual differences in detection thresholds for vibrotactile stimuli and the variations in tactor characteristics, it is important to calibrate the perceived intensity of the 39 haptic symbols across participants and equalize the perceived intensity of the 24 tactors. This is achieved with a two-step calibration procedure. First, detection thresholds at 60 and 300 Hz were estimated with one tactor (i.e., the "reference tactor" in Row ii, Column 4, Figure 1) using a one-up two-down adaptive procedure [32]. Second, the intensities of all 24 tactors were adjusted to match that of the reference tactor using a method of adjustment [32]. The two-step calibration was performed for each participant prior to the three studies reported in this article. A detailed description of the calibration procedures can be found in [30].

III. STUDY I: 10 PHONEMES AND 51 WORDS

The first study examined the feasibility of learning 10 of the 39 phonemes and 51 English words made up of the 10 phonemes.

A. Methods

1) Participants: Ten naïve participants (P01 to P10, 5 females; age range 18-30 years old, average 21.6 ± 3.4 years old) took part in Study I. All were right handed with no known sensory or motor impairments. The participants came from diverse language backgrounds. All participants were fluent in English, including four native English speakers. Other languages spoken among the participants included Arabic, Bhojpuri, Bulgarian, Chinese, French, German, Hindi, Japanese, Korean, Maithili, Nepali, Punjabi, Tharu, and Tibetan. Most of the participants also received early childhood music training including piano, sax, clarinet, percussion, pipa, trumpet, violin and zither.

Experimental data from participants P01-P04 were obtained and reported in a previous study [33].

2) Learning Materials: The learning materials consisted of 10 phonemes and 51 English words. The phonemes were: EE, AY, OO, I, D, M, S, W, DH, and K. Table III lists the 51 words made up of the 10 phonemes divided into two groups. The full set of 51 words consist of 23 VC (vowel-consonant) or CV (consonant-vowel) words and 28 CVC (consonant-vowel-consonant) words. Each word was transcribed into the corresponding English phonemes and presented as a concatenated

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

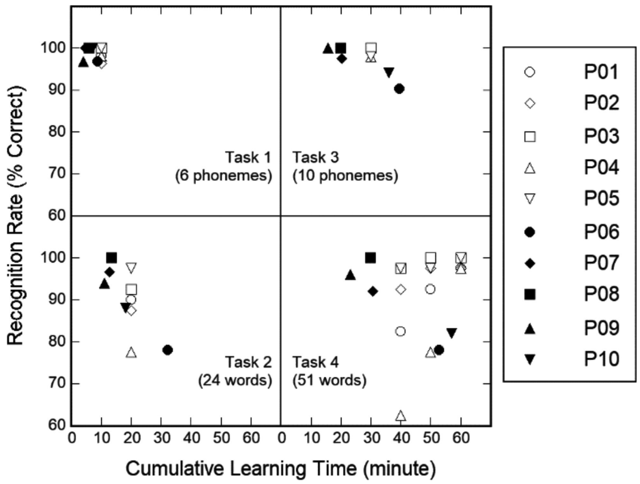

Fig. 3. Results of Study I on the learning of 10 phonemes and 51 words for the ten participants. Each panel shows the results for one of the four tasks. The open symbols represent participants P01 to P05 who followed a timed learning procedure. The filled symbols represent participants P06 to P10 who followed a self-paced learning procedure.

sequence of the corresponding haptic symbols, with an inter-phoneme interval of 300 ms between adjacent phonemes. The haptic symbols for the 10 phonemes used in Study I were as described in Tables I and II and Figures 2 and 3 of [30], with the following two exceptions. The haptic symbols for "D" and "S" used an earlier version of the codebook in which the waveform for "D" was 60 Hz and the location of "S" was on the volar mid-forearm.

3) Learning and Testing Procedures: The learning of the 10 phonemes and 51 English words were divided into four tasks that got progressively more difficult. They were:

- Task 1: 6 phonemes (AY, S, D, OO, M, EE);
- Task 2: 24 words (Group 1) made up of these 6 phonemes;
- Task 3: 10 phonemes (4 new added to Task 1: I, W, DH, K);
- Task 4: 51 words (Group 1 + Group 2) made up of the 10 phonemes.

Each learning task consisted of "free play" and "practice identification test." During the free play, the participant selected a phoneme or word to be learned, and felt the haptic symbol(s). During a practice identification test, the participant felt a phoneme or word, gave a response, and received trial-by-trial correct-answer feedback. Time spent in both learning activities was logged as learning time. Additional testing was conducted without any feedback to gauge the participant's performance level for the self-paced group (see below). Time spent on the tests without feedback was not included in the calculation of cumulative learning time. The identification tests for phonemes and words employed closed-set testing where the complete list of stimuli for a given task was made available to the participants, from which they were instructed to select a response on each trial. The layout of the phonemes or words on the computer screen was the same for learning and testing in each task.

To compare the performance with a timed vs. self-paced procedure, the ten participants were randomly assigned to two groups. For participants P01-P05, the learning time was limited to 10 minutes on each learning day, for a total of six days. Each of the six learning days included 5 minutes of "free play" followed by 5 minutes of "practice identification test." No test without feedback was conducted for the participants in this group. The short time period allowed for learning on each day helped the participants to maintain full concentration during the time spent learning the phonemes and words. By spreading the 60-min learning time over six days, we sought to take advantage of memory consolidation; i.e., improvement in phoneme and word recognition performance after a period of time when the participant was not actively engaged in the learning task [34]. Evidence for memory consolidation during the learning of all 39 phonemes was reported earlier in [33] (see Section 6, Exp. II).

The learning activities for participants P01-P05 were organized as follows:

- Day 1: Task 1 (6 phonemes learned and tested);
- Day 2: Task 2 (Group 1 words learned and tested);
- Day 3: Task 3 (4 new phonemes learned, 10 tested);
- Day 4: Task 4 (Group 2 words learned, 51 words tested);
- Day 5: Task 4 (51 words reviewed and tested);
- Day 6: Task 4 (51 words reviewed and tested).

The number of trials for the tests was as follows: 54 trials for 6 phonemes (Task 1), 40 trials for 24 words (Task 2), 50 trials for 10 phonemes (Task 3), and 40 trials for 51 words.

The remaining participants (P06-10) followed a self-paced learning procedure conducted within one or two 2-hour laboratory sessions. They were trained and tested on the four tasks described above. For each of the four tasks, the self-paced training began with the use of "free play," followed by a "practice identification test" with trial-by-trial correct-answer feedback as described above. When a criterion level of performance $\geq$ 80%-correct was achieved on a practice test, the participant was tested without feedback, and then advanced to the next task on the list.

On the phoneme identification task, 24 trials were presented in each of tests with and without feedback for Task 1 and 32 trials for Task 3. For the word identification testing, 30 trials were employed in each run of testing with feedback and 50 trials in the non-feedback testing for both Task 2 and Task 4.

B. Results and Discussion

The percent-correct scores for phoneme recognition (Task 1 and 3) and word recognition (Task 2 and 4) from the identification tests without feedback are shown in Figure 3 as a function of cumulative learning time in minutes. For Task 1 (learning of 6 phonemes; see the upper-left panel of Figure 3), all ten participants reached a phoneme recognition rate of 96.3 to 100% correct within the first 10 minutes of cumulative learning time, with the self-paced group (P06 to P10, filled symbols) doing so within 4.15 to 8.81 minutes. For Task 2 (learning of 24 words), the participants reached a word recognition rate of 77.5 to 100% correct

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
THE ONE HUNDRED ENGLISH WORDS USED IN STUDY II

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8
ace	key	ray	low	chow	jay	gay	all
aid	my	she	oath	cow	joy	go	fee
day	sigh	shy	row	how	knee	guy	off
may	they	shoe	show	vow	no	tie	on
me	way	us	so	wow	now	toe	ought
moo	why	come	the	cheese	pay	too	you
sue	woo	rock	though	choose	join	toy	fought
doom	dime	rum	base	hatch	keep	azure	pawn
dude	make	shave	dome	him	noise	book	ring
moose	seek	shock	like	loud	pen	gun	thing
same	side	vase	thumb	mad	them	put	young
seam	wake	wash	will	maze	then	shirt	your
seed	weed		wish			should	

within 32.33 minutes of cumulative learning time. For Task 3 (learning of 4 new phonemes and testing with 10 phonemes), the participants reached 90.3 to 100% correct within 39.64 minutes. There are more data shown in Figure 3 for Task 4 (see the lower-right panel) because participants P01 to P05 did the task from Day 4 to Day 6. These participants in the timed learning procedures were able to learn all 51 words with a recognition rate of 97.5 to 100% correct by the end of Day 6 (60 minutes). With the self-paced learning procedure, participants P06 to P10 reached 78.0 to 100% within 56.98 minutes. Specifically, three of the participants (P07, P08, and P09) achieved criterion performance on the 51-word task within one 2-hour session. One participant (P06) achieved 78.0% correct which was considered close enough to the 80% criterion. The remaining participant (P10) required a second session to achieve the goals of the study. There does not appear to be any systematic difference in the performance level and learning time between the two participant groups.

It was relatively easy for the participants to learn the 10 phonemes. Phoneme recognition rate was nearly perfect for Task 1 (6 phonemes) and above 90% for Task 3 (10 phonemes). For some participants, the transition from phonemes to words required some familiarization, as seen comparing the results from Task 1 (6 phonemes) and Task 2 (24 words made up of the 6 phonemes). This indicates that additional learning was needed to process phonemes delivered in a sequence. Despite the initial "dip" in performance for Task 2 and Task 4 when the participants transitioned from phonemes to words, word-recognition improved quickly for Task 4 as seen in the large performance improvement by P01 (82.5 to 97.5% correct) and P04 (62.5 to 97.5% correct) from 40 to 60 minutes. With additional learning time, participants P06 and P10 might have improved their Task 4 performance levels significantly as well.

Despite some individual differences, all participants succeeded in learning the 51 words with very few errors within one hour, which is a very reasonable amount of cumulative learning time. The average of the cumulative learning time for all participants (using 60 minutes for P01-P05) was 49.33 minutes. It corresponds to an average learning rate of *1.0 word per minute*.

IV. STUDY II: 39 PHONEMES AND 100 WORDS

Encouraged by the high performance level achieved in Study I within one hour of learning, the second study expanded the phoneme list to include all 39 English phonemes and increased the word list to 100 words made up of all the phonemes. The objectives of Study II were threefold. First, it was important that we assess the learning of *all* English phonemes, so that *any* English word can potentially be presented and acquired on the skin. Second, the learning time should be manageable so that our tactile speech communication system has a practical impact. Third, we compared the relative merits of phoneme-based (bottom-up) and word-based (top-down) learning approaches to find the most efficient learning paradigm. The results of Study II were reported earlier in [35] and re-analyzed here.

A. Methods

1) Participants: A total of twenty-four new naïve participants took part in Study II. Twelve of them (P11-P22; 6 females; age range 18-26 years old, average age 21.9 ± 1.7 years old) were randomly assigned to the phoneme-based learning group. The remaining twelve (P23-P34; 6 females; age range 19-39 years old, average age 25.0 ± 5.7 years old) were assigned to the word-based learning group. All were right handed with no known sensory or motor impairments. Six of the participants in each group are native English speakers. The other participants speak English fluently and their first languages include Bulgarian, Chinese and Korean. Most of the participants received early childhood music training including piano, violin, guitar, flute, and cello.

2) Learning Materials: The learning materials consisted of 39 phonemes and 100 common English words. The haptic symbols for the 39 phonemes were as described in Tables I and II and Figures 2 and 3 of [30]. The words were organized into eight groups (see Table IV; the groups are explained later in Section 4.1.5). They consisted of 50 two-phoneme (CV or VC) words and 50 three-phoneme (49 CVC and 1 VCV) words. Each word was transcribed into the corresponding English phonemes. Each phoneme was mapped to the corresponding haptic

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

symbol, with an inter-phoneme interval of 300 ms between phonemes. The word duration varied from roughly 1 to 2 s.

3) General Learning and Testing Procedures: For both phoneme-based and word-based groups, the learning procedure was based on a 10-day curriculum where learning time was capped at 10 minutes per day, for a total of 100 minutes. On each day, the participants engaged in free play and practice identification testing, and the total time was recorded as learning time. The participants were encouraged to spend time with both activities during the 10-min learning period, and could decide how to divide their time between them. In order to assess learning progress, a closed-set phoneme or word identification test without feedback was conducted after the 10-min learning period. The test typically took less than 10 minutes and did not count towards learning time since no correct-answer feedback was provided. The procedure was followed by the participants in both the phoneme-based and word-based learning groups, so their results could be compared under comparable conditions.

The combined experimental time, excluding the pre-experiment threshold testing and tactor intensity calibration time, reached 80 hours (24 participants × 10 days × 1/3 hour per day). The procedures followed by the two groups of participants are outlined below for phoneme-based and word-based learning, respectively.

4) Procedure for Phoneme-Based Learning: The 10-day curriculum for phoneme-based learning was as follows:

- Day 1: 6 Cs (consonants) – P T K B D G;
- Day 2: 12 Cs – Day 1 + F V TH DH S Z;
- Day 3: 18 Cs – Day 2 + SH ZH CH J H W;
- Day 4: all 24 Cs – Day 3 + M N NG R L Y;
- Day 5: 8 Vs (vowels) – EE IH AH OO UU AE AW ER;
- Day 6: 15 Vs – Day 5 + AY I OW OE OY UH EH;
- Day 7: all 39 phonemes (> 90% correct required before learning words);
- Day 8: 50 VC/CV words (if > 90% correct with 39 phonemes, otherwise repeat Day 7);
- Day 9 & 10: all 100 words (after one day with 50 VC/CV words);

With the phoneme-based learning approach, participants P11-P22 learned the haptic symbols associated with the 39 phonemes before learning the 100 words presented as sequences of phonemes. As shown above, the 24 consonants were divided evenly into 4 groups and learned from Days 1 to 4. The 15 vowels were divided into two groups and learned during Days 5 and 6. On Day 7, all 39 phonemes were available for learning and each participant had to achieve at least 90% correct on a phoneme identification test before proceeding to learning words. Therefore, all 12 participants had the same learning tasks from Day 1 to 7.

From Day 8, the participants who had successfully passed the 90%-correct phoneme identification criterion spent their 10-min learning time on free play and practice identification test, this time with words instead of phonemes. Again, the participant completed a word identification test without any feedback after the 10-min learning period was over. The 100 words were divided into two groups: the first 50 words consisting of only two-phoneme words and the remaining 50 consisting of three-phoneme words. After reaching the 90% criterion for phoneme learning, each participant learned the 50 CV/VC words for one day only regardless of their performance level. This was followed by all 100 words on the following day until 10 days were reached. It follows that the participants may proceed at different paces from Day 8 to Day 10 due to differences in individual progress.

5) Procedure for Word-Based Learning: With the word-based learning approach, participants P23-P34 started with word learning on Day 1. To gradually increase the difficulty levels, the 100 words were divided into 8 groups with an increasing number of phonemes contained in each group (see Table IV). For example, the 13 words in Group 1 were made up of 6 phonemes: D, M, S, AY, EE and OO. Each successive group contained 12 to 13 additional words with 4 to 5 additional phonemes, as shown below.

- Group 1: 13 words (6 phonemes);
- Group 2: 13 words (4 new phonemes);
- Group 3: 12 words (5 new phonemes);
- Group 4: 13 words (5 new phonemes);
- Group 5: 12 words (5 new phonemes);
- Group 6: 12 words (5 new phonemes);
- Group 7: 13 words (5 new phonemes);
- Group 8: 12 words (4 new phonemes).

A performance level of 80% correct had to be reached before a participant could proceed to the next group of words on the following day. At the end of each day, the participant was tested with all the words s/he had learned so far. The process continued until 10 learning days were completed. Participants who reached Group 8 before Day 10 continued with all 100 words until Day 10. As an example, a participant who succeeded in passing the performance criterion at the end of each day would be tested with 13, 26, 38, 51, 63, 75, 88, 100, 100, and 100 words from Days 1 to 10, respectively.

B. Results and Discussion

The results in terms of learning time and performance levels are presented below, first for the phoneme-based and word-based groups separately, then combined for comparison. By design, different tasks were performed on different learning days, with some tasks much easier than others. Instead of reporting the percent-correct scores which are task dependent, we report the "equivalent number of words learned (ENW)" by multiplying the percent-correct scores (PC) with the corresponding number of words (NW) in the closed stimulus set used in tests conducted at the end of each day without any feedback.

$$ENW = PC \times NW$$

For phoneme-based learning, we also calculate and report the "equivalent number of consonants/vowels/phonemes learned."

1) Results of Phoneme-Based Learning: The participants in the phoneme-based learning group all performed the same tasks from Day 1 to Day 7. The equivalent number of consonants,

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

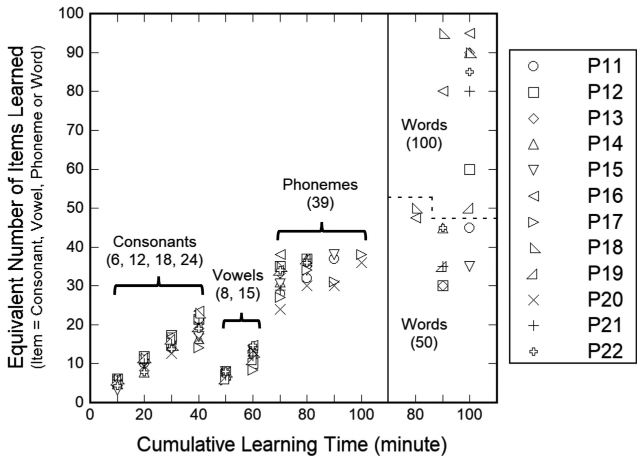

Fig. 4. Results of phoneme-based learning in Study II: Equivalent number of items learned as a function of cumulative learning time. The left panel shows the learning of consonants, vowels and all phonemes. All participants performed the same task up to Day 7 (70 minutes). Those who did not meet the >90% correct criterion continued with all phonemes on Day 8, etc. The right panel shows the progress for word learning. As an example, if a symbol starts at 90 minutes (e.g., □), it means that the corresponding participant (i.e., P12) achieved the phoneme performance criterion on Day 8 and proceeded to word learning on Day 9. The data points for 50 words and 100 words are separated by the dashed line drawn around 50 items in the middle of the right panel.

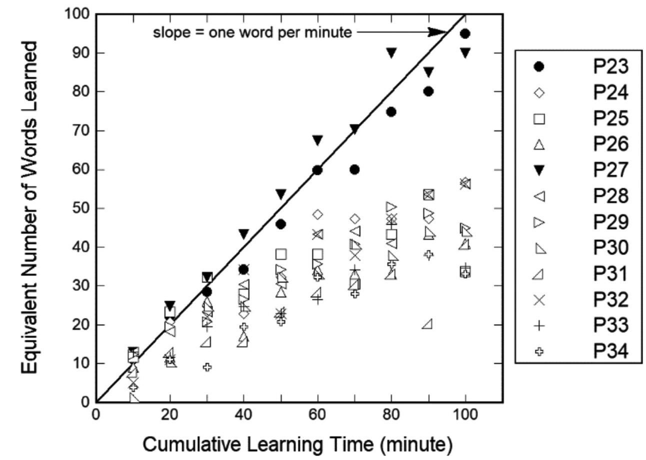

Fig. 5. Results of word-based learning in Study II: Equivalent number of words learned as a function of cumulative learning time. The solid line corresponds to a learning rate of one word per minute.

vowels and phonemes learned are shown in the left panel of Figure 4. The numbers under "Consonants," "Vowels," and "Phonemes" indicate the total number of items tested at the end of each 10-minute learning period. For example, six consonants were learned and tested at the end of 10 minutes of learning (Day 1), and 15 vowels were tested at 60 minutes (Day 6). In order for the participants to meet criterion $>$ 90% correct with 39 phonemes on Day 7, the number of items learned needed to be at least 35 phonemes. Only two participants (P16, P18) succeeded at 70 minutes. They proceeded to practice with 50 CV/VC words on Day 8 and learned 47 and 50 words, respectively (see the right panel of Figure 4 at 80 minutes). Both participants practiced with 100 words on Days 9 and 10, and learned 90 to 95 words by the end of the 10-day period. From the left panel of Figure 4, it can be seen that the remaining ten participants continued with 39 phonemes on Day 8. Six of the ten participants (P12, P13, P14, P19, P21 and P22) passed the performance criterion, practiced with 50 words on Day 9 (right panel), and tried all 100 words on Day 10. They learned between 50 to 90 words by 100 minutes. Among the remaining four participants, P11 and P15 learned 37 and 38 phonemes, respectively, at 90 minutes (left panel), and practiced with 50 words on Day 10 (right panel). The remaining two participants (P17 and P20) reached the 39-phoneme performance criterion by Day 10 (left panel at 100 minutes), but never tried any words.

The results obtained with the phoneme-based learning approach demonstrate that all the participants were able to learn the haptic symbols associated with the 39 English phonemes with a $>$ 90% accuracy within 100 minutes. Individual learning outcomes varied, and half of the twelve participants were able to learn the 100 English words with scores $>$ 80% correct by the end of the 10-day learning period.

2) Results of Word-Based Learning: Figure 5 shows the equivalent number of words learned for the word-based learning group. Recall that the participants in this group were required to reach $>$ 80% correct when tested with the cumulative groups of words practiced so far before a new word group could be added to the learning. Of the twelve participants, 2 participants (P23 and P27; filled symbols in Figure 5) were able to reach the criterion with all 8 groups of words (totaling 100 words) and learned 95 and 90 words, respectively, by the end of 100 minutes. Five participants (P24, P25, P28, P29, P32) reached the criterion with the first 5 groups of words (63 words), 4 participants (P26, P30, P31, P33) with 4 groups (51 words), and 1 participant (P34) with 3 groups (38 words). It thus appears that there is a large performance gap between the top 2 participants and the remaining 10 participants. This is clearly observable in Figure 5 where the learning rate for P23 and P27 remained roughly one word per minute throughout the course of the study, but rate of learning for the remaining participants decreased as time went on. The data points for the top two performers and the rest of the participants started to diverge after 30 minutes of learning. The performance of the 10 participants plateaued at 57 or fewer words and averaged 44 words (about half the number of words learned by P23 and P27) at the end of 100 minutes.

3) Comparison of Phoneme-Based vs. Word-Based Learning: The performance comparison between the phoneme-based and word-based learning groups focuses on the word learning tasks using the common metric "equivalent number of words learned" (Figure 6). Recall that the participants in the phoneme-based learning experiment did not reach word learning until Day 8 or later. Thus, the data from the two groups are replotted for the final 30 minutes (Day 8 to Day 10) for ten of the participants in the phoneme-based learning group (excluding P17 and P20 who never reached word learning) and all twelve participants in the word-based group. For the participants in the phoneme-based learning group (left panel of Figure 6), two participants (P16, P18) learned 47.5

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

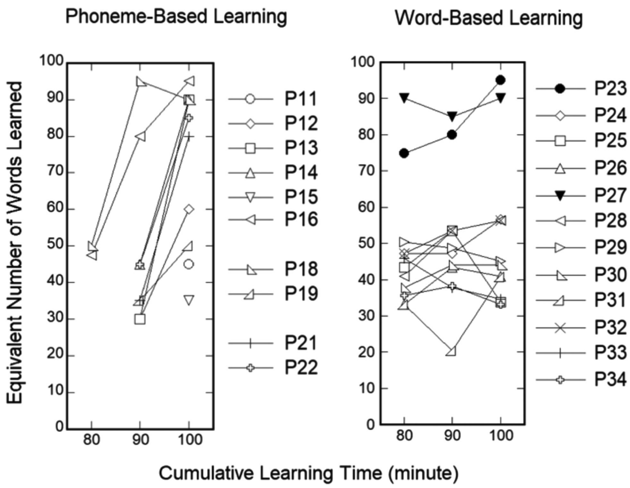

Fig. 6. Comparison of equivalent number of words learned on Days 8, 9, and 10 by participants in the phoneme-based learning group (left panel) and word-based learning group (right panel). The data points for the same participant are connected to show rate of improvement.

and 50 words on Day 8, jumped to 80 and 95 words on Day 9, and ended at 95 and 90 words on Day 10, respectively. Six participants joined word learning on Day 9 and two more on Day 10. The lines demonstrate a clear upward trend for each participant, with the four participants in the middle of the performance range (P13, P14, P21 and P22) showing the largest improvement from Day 9 to Day 10. Although to a lesser extent, the two remaining participants (P12 and P19) clearly improved from Day 9 to 10. It appears conceivable that given more learning time, the participants in the phoneme-based learning group would continue to improve and eventually learn the 100 English words with an error rate $\leq 20\%$.

Data for the participants in the word-based group (right panel of Figure 6) show a different pattern. There are clearly two groups of learners, with the 2 participants P23 and P27 (filled symbols) significantly outperforming the remaining 10 participants. Furthermore, the 10 participants appear to be reaching plateaus from Day 8 to Day 10, leaving it unclear whether these 10 participants in the lower performing group would ever reach 100 words. During post-experiment debriefing, the two top performers attributed their success to learning the phoneme codes early in the learning process. The other participants took longer to realize that words were composed of phoneme sequences. Several possible reasons may account for the plateau observed in their performance. For example, they may have only learned a limited number of phonemes, thus reducing their ability to identify the words. Or, if their strategy had been to memorize words as a whole, they were likely limited in the number of words that could be acquired in this manner.

Focusing on the last day of performance for both learning approaches (see data points at 100 minutes in both panels of Figure 6), there is a large spread of words learned among the participants in each group. Both phoneme-based and word-based learning approaches appear to be feasible for the acquisition of 100 English words in that there are examples of high-performing participants who have acquired at least 90 words in either group. However, the performance distributions for the two groups suggest that the phoneme-based approach leads to a consistent pattern of improvement, with half of the 12 participants acquiring at least 80 words in 100 minutes. In comparison, only 2 of the 12 participants in the word-based learning group achieved the same. The comparatively poorer performance of the word-based group may be considered in light of the phoneme-based coding strategy taken in constructing the words, which may have introduced a bias in favor of the participants who were introduced to the phonemes first.

V. Study III: 39 Phonemes and 500 Words

The objectives of Study III were to demonstrate that there was no limit to the vocabulary size and that the same learning rate could be maintained. Due to the need for manual transcription of words into phoneme strings and for performance testing, a closed word list containing 500 English words was constructed (see supplemental materials). The vocabulary size was large enough that this study can be effectively considered as an open-vocabulary study.

There were two parts to Study III. First, a group of experienced participants who had already acquired the 39 phonemes and 100 English words on TAPS learned the expanded word list of 500 words. This "generalizability" experiment allowed us to gauge the feasibility of learning a much larger vocabulary while leveraging the hours already spent on acquiring 39 phonemes and 100 words. Upon the encouraging results obtained with the experienced participants, an additional group of naïve participants was recruited to acquire all 39 phonemes and 500 words from scratch, so that their learning time could be documented through the entire process.

A. Methods

1) Participants: A total of twenty-one participants took part in Study III. Eleven of them (P05, P18, P23, P27, P35-P41) had already learned the 39 phonemes and 100 words prior to Study III. They included one participant (P05) who was in Study I, three participants (P18, P23, P27) who were in Study II, and seven participants (P35-P41) who did not take part in either Study I or Study II. Participants P05 and P35-P41 learned the 39 phonemes and 100 words in other earlier studies that are not reported here. The remaining ten (P42-P51) were naïve participants who had not been exposed to TAPS before. Among the twenty-one participants, there were 10 females, and an age range of 18 to 27 years old, averaging 21.5 ± 1.9 years. All participants were right handed with no known sensory or motor impairments. Fifteen of the participants are native English speakers, and the remainder are fluent in spoken English. Other languages spoken by the participants include Arabic, Chinese, French, German, Korean, Portuguese, Spanish, and Swedish. Most of the participants received early childhood music training including piano, cello, clarinet, flute, guitar, violin, and trumpet.

2) Learning Materials: The learning materials consisted of 39 phonemes (for the naïve participants only) and 500 common English words (for both the experienced and naïve

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

TABLE V
THE TEN PHONEME AND WORD GROUPS FOR THE NAÏVE PARTICIPANTS IN STUDY III

Group	1	2	3	4	5	6	7	8	9	10
New Phonemes	D S M AY EE OO	DH K W I	SH V R AH UH	B L TH OE IH	CH Z H AE OW	P N J OY EH	T G ZH ER UU	F Y NG AW	-	-
Number of Phonemes	6	10	15	20	25	30	35	39	39	39
Criterion (Phonemes)	> 80%	> 80%	> 80%	> 80%	> 80%	> 80%	> 80%	> 80%	-	-
Number of Words	24	51	76	101	126	151	176	201	251	500
Criterion (Words)	50-60%	50-60%	50-60%	75-85%	75-85%	75-85%	75-85%	> 80%	> 70%	-

participants). The tactile phonemic codes used in the current study were as described in Tables I and II and Figures 2 and 3 of [30], with the following exceptions. The durations of the six plosive phonemes (P, B, T, D, K, G in Table I and Fig. 2 of [30]) were increased from 100 to 140 ms, and the durations of the 11 vowel and diphthong stimuli that were previously 480 ms were decreased to 400 ms. The changes in signal durations made the plosives easier to perceive and did not significantly impact the performance with the vowels and diphthongs. The experienced participants showed no difficulty in adjusting to the modified signal durations. For the naïve participants, the phonemes and words were organized into ten groups (see Table V). The 39 phonemes were divided into eight groups, each containing 4-6 phonemes. Once the participant reached the performance criterion of 80% or higher for the phoneme group, the next group of new phonemes was added. This way, the phonemes were learned in a cumulative way and the number of phonemes in each group grew from 6 to 39 phonemes from Group 1 to Group 8. Concurrent with phoneme learning, words made up of the phonemes that participants had learned up to that point were introduced, from 24 to 201 words from Group 1 to Group 8, 251 words in Group 9, and 500 words in Group 10 (see Table V). The words in Groups 1 through 8 consisted of 2-phoneme (CV or VC) and 3-phoneme words (primarily CVC). Of the 201 words in Group 8, 63 were 2-phoneme words and 138 were 3-phoneme words. For the Group 9 set, an additional 50 words were added to Group 8: 10 2-phoneme words and 40 3-phoneme words. Of the full set of 500 words in Group 10, 2 words had 1 phoneme, 89 words had 2 phonemes, 359 words had 3 phonemes, 49 words had 4 phonemes, and 1 word had 5 phonemes. In addition to CV, VC, and CVC structure, words in the 500-word set also included consonant blends, as in VCC constructions (e.g., ask), CCVC (e.g., glad), and CVCC (e.g., coast).

The experienced participants started with a review of the 100 words they had already learned, and proceeded with word lists consisting of 150, 200, 250 and 500 words.

The inter-phoneme interval was reduced from 300 ms to 150 ms for all the words. This corresponded to word presentation rates of roughly 36 words/min assuming a 500-ms interval between words.

3) Learning and Testing Procedures: The design of the learning curriculum for the naïve participants took into account the insights gained from Studies I and II. A major finding of Study II was that phoneme-based learning led to a continuous improvement in performance. Therefore, the naïve participants in Study III spent time learning the individual phonemes prior to and during the learning of English words. A key finding of Study I was the temporary drop in performance when the participants switched from phoneme to word learning. In Study III, the naïve participants practiced with a group of words composed from the set of the cumulative number of phonemes they had been learned up to that point, and the number of words in the list grew with the number of phonemes learned. It was hoped that the phoneme-based learning would build a solid foundation for word learning, and that the mix of phonemes and words during the learning process would introduce the participants to the reception of multiple phonemes making up a word early in the learning process.

The procedures for the naïve participants were more consistent than those for the experienced participants, and are described here first. The participants spent up to two hours per day over a three-to-four-week period in learning and testing, except for P51 whose sessions were spread out over seven weeks due to scheduling issues. For each group of phonemes and words, the participants began with free play. They could select any phoneme or word to be practiced, and either feel the stimulus on TAPS or look at a visual representation of the tactor activation sequence. When ready, the participants proceeded to practice identification tests for phoneme or word identification with correct-answer feedback. They were allowed to re-play any stimulus as they wished. The total time spent on free play and practice identification tests was logged as learning time. At the end of each day, a phoneme or word identification test was conducted to gauge the participant's performance level. The performance criteria as specified in Table V had to be met for a group for both the phonemes and words before a participant could move on to the next group.

Within each group, the participants always learned the phonemes before the words. When words were tested, the full list of words was shown on the screen up to Group 4 (101 words),

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

and the participants could select one of the displayed words as the response. For Group 5 and higher where the number of words in the list exceeded 101 words, the list was no longer shown on the screen. The participants had to type a word into a text box as a response. The migration from closed-set responses to open-set responses was important for the assessment of performance in a real-word communication scenario.

The performance criterion for phonemes was 80% or higher for Groups 1 to 8. We found from previous studies that this percent-correct level demonstrated an adequate proficiency and allowed the participants to continue to improve with the subsequent groups. The performance criterion for words was set to 50-60% or higher initially for Groups 1 to 3, when the participants were still getting used to acquiring multiple phonemes at a time. The criterion was raised to 75-85% for Groups 4-7, 80% for Group 8 (201 words), and dropped slightly to 70% for Group 9 (251 words). The participants aimed for 70% or higher performance with Group 10 (all 500 words). If this criterion was not achieved within four 50-trial runs with feedback, they were allowed to proceed to the testing phase. At the end of the experiment, each participant was tested with 10 runs of 50 trials for a total of 500 trials with open-set responses and without feedback. For each of the 10 runs, 50 words were randomly selected from the 500 word list (i.e., randomization with replacement).

The procedures for the experienced participants varied in the way that phonemes were acquired. Of the eleven experienced participants, three (P18, P37-P38) learned all 39 phonemes prior to learning any words. Six participants (P05, P35-P36, P39-P41) followed a learning procedure similar to that of the naïve participants outlined above, and learned the 39 phonemes and some words together. The remaining two participants (P23, P27) acquired words directly without learning phonemes individually. Most of the experienced participants practiced and were tested with word lists containing 100, 150, 200, 250 and 500 words, except for four participants (P35-P36, P40-P41) who were only tested at 250 and 500 words. During training, the participants identified words selected at random from a given word list. Open-set responses were entered into a text box. On error trials, the correct word appeared as written text on the computer screen. During testing, the participants performed word identification without feedback. Like the naïve participants, the experienced participants were tested on the 500-word vocabulary with 10 runs of 50 trials at the end of the experiment with open-set responses and without any feedback.

B. Results and Discussion

1) Results of Experienced Participants: Performance levels of the experienced participants are converted to "Equivalent Number of Words Learned" by multiplying the percent-correct score by the number of words in the closed vocabulary list, and are plotted as a function of additional learning time in minutes after the acquisition of 100 words (see Figure 7). Each of the seven participants shown with open symbols has five data points that correspond to the equivalent number of words learned with a word list containing 100, 150, 200, 250

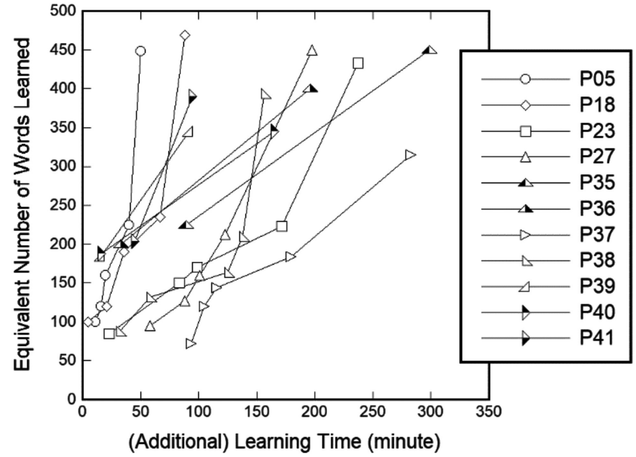

Fig. 7. Word identification performance of experienced participants in Study III after they had acquired the initial 100 English words. Shown are equivalent number of words learned as a function of the additional learning time.

and 500 words, respectively. Each of the four participants shown with half-filled triangles has only two data points for 250 and 500 words, respectively. Some inter-subject variability can be observed in Figure 7. The additional learning time (beyond the first 100 words) required to reach the 500-word vocabulary ranged from 50 min (less than an hour) to 299 min (5 hours). The 500-word test scores ranged from 63% to 94% correct across the eleven experienced participants, with eight scoring higher than 78% correct. Across all the experienced participants, the highest equivalent number of words learned ranged from 315 words (P37) to 469 words (P18). In terms of learning time, the best participant P18 took the second shortest additional time (88 min) and the worst participant (P37) took the second longest additional time (283 min).

The results with the experienced participants demonstrated the capability of open-set word identification with good accuracy within a reasonably short period of learning time measured in hours rather than months. The results proved the 39 haptic symbols used to represent the 39 English phonemes to be adequate. Furthermore, the data trend indicated that word recognition accuracy would continue to improve with further practice, and performance could be reasonably maintained with further increases in vocabulary size. The promising results obtained with the experienced participants encouraged us to proceed with the group of naïve participants (P42-P51) to (1) test their ability to acquire 500 English words from scratch, and (2) document the total learning time needed to acquire 500 words.

2) Results of Naïve Participants: The phoneme identification performance of the naïve participants is shown in Figure 8. Phoneme identification scores were in excess of 85% correct for most participants within each phoneme group. With the full set of 39 phonemes in Group 8, 9 of the 10 naïve participants scored above 90% (92%-99%) correct, and one participant (P46) reached 84% correct. Although P46's performance was sufficient for moving onto the next group, this participant consistently scored the lowest with word recognition, as shown in Figure 9.

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

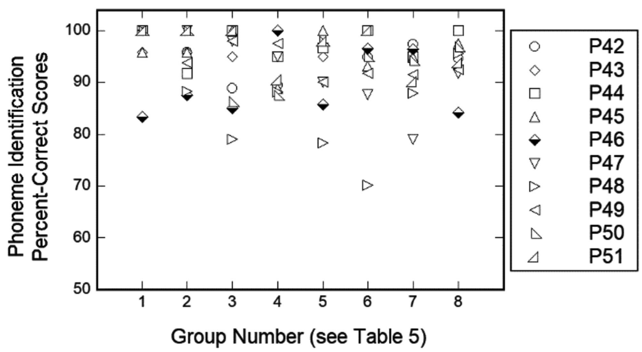

Fig. 8. Percent-correct scores for phoneme identification by the naïve participants in Study III. The number of phonemes in each group was 6, 10, 15, 20, 25, 30, 35, and 39 from Group 1 to 8. See Table V for the phonemes included in each group.

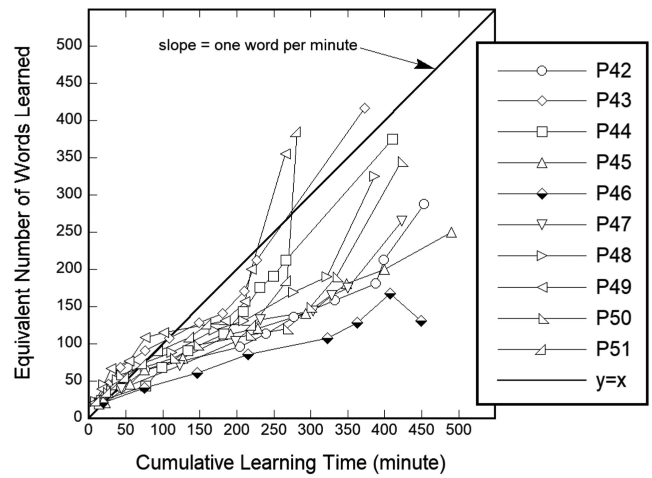

Fig. 9. Performance with word identification by the naïve participants in Study III. Shown are the equivalent number of words learned as a function of the cumulative learning time.

The results of phoneme identification indicated that the haptic symbols were sufficient for conveying information about the 39 phonemes and were well recognized when presented in isolation.

Figure 9 shows the naïve participants' word identification performance as a function of the cumulative learning time for each individual participant. The percent-correct scores fluctuated for each participant. We again calculated the equivalent number of words learned by multiplying the percent-correct scores by the number of words in the word list, for each of the ten groups shown in Table V. The cumulative learning time included time spent in "free play" and "practice identification tests with feedback" on the phoneme and word groups up to the current word group. It follows that there are 10 data points per participant corresponding successively to an increasing vocabulary size of 24, 51, 76, 101, 126, 151, 176, 201, 250 and 500 words. The only exception is the 8 data points for participant P46. This participant had a relatively lower phoneme identification accuracy with 39 phonemes in Group 8 (see half-filled diamond symbols in Figure 8), and was consistently ranked the lowest with word identification (see half-filled diamond symbols in Figure 9). It typically took 2-3 days for P46 to reach the performance criterion with each group. By the time the three-week experimental period was over, this participant had only reached a vocabulary size of 201 words (Group 8) and scored 65% correct.

The solid straight line in Figure 9 corresponds to a learning rate of one word per minute. Within the first 100 minutes of learning, the learning rate among the naïve participants clustered around the reference line of one word per minute. The learning rate then dropped for cumulative learning time between 100 and 200 minutes as the number of phonemes and words increased with group number. Individual differences among the participants became more pronounced after the first 200 minutes of learning. The performance of some participants (e.g., P49, P51) shot up rapidly while that of others (e.g., P42, P45) continued at a slower pace. The participants can be grouped into three categories. The six participants in the top group (P43-44, P48-51) acquired 325 to 417 words within 266 to 423 minutes (about 4.5 to 7 hours) of cumulative learning time. The three participants in the middle group (P42, P45, P47) acquired 250-288 words within 7-8 hours of learning. The bottom participant (P46) advanced only to Group 8 and acquired 131 words after 7.5 hours of learning. Generally speaking, the learning curves were steeper between 250 to 500 words than for the first 250 words. For the top performing participants on the 500-word test, the average word acquisition rates were roughly 1.3 words/min. This was slightly higher than the rates achieved by the top-performing participants in the word-based learning group in Study II (see Figure 5).

VI. General Discussion

We have reported three studies that assessed people's ability to acquire English words through a vibrotactile display that encoded the 39 English phonemes and delivered them as building blocks on the forearm. We have developed a learning procedure that involved the gradual introduction of the 39 phonemes in 4-6 phoneme groups, and mixing phoneme and word learning early on to facilitate the reception of phoneme streams. Learning was spread over multiple sessions and an entire experiment lasted several weeks for each participant. Data from 51 participants (4 of them participated in more than one study) are reported in this article. Forty-one of the participants learned all 39 phonemes and at least 100 English words. Twenty-one participants learned up to 500 words and used open-set responses by typing words during word identification tests. The cumulative learning time for each participant was up to 8 hours, rather than weeks or months as has been reported for other tactile word learning studies [36]– [40]. The results provide ample evidence that any English word can be transmitted through our phonemic-based tactile speech communication system TAPS, learning occurs in small chunks of time, and the total learning time is reasonable.

Our results can be compared to those from several recent projects aimed at transmitting English speech through the skin. Table VI shows the studies in two groups and compares the key parameters of all studies. The first group consists of three studies: two conducted at Facebook using tactile arrays

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
COMPARISON OF RECENT STUDIES OF TACTILE SPEECH COMMUNICATION SYSTEMS (SEE TEXT FOR ENTRIES MARKED WITH *)

Study Author (Year)	# Tactors	Learning Time	# Symbols	# Words	Open/Closed Responses	Chance Level	Word Scores
Zhao et al (2018)	6	30 min	9 phonemes	20	Open-set	< 5%	83%
Turcott et al. (2018)	16	50 min	10 phonemes	20	Closed-set	5%	76%
Dunkelberger et al. (2018)	4, squeeze, skin stretch	100 min	23 phonemes	150/50	Closed-set (*self-paced*)	2%	87%
Novich (2015)	27	11-12 days 300 trials/day	Spectral-based	50	Closed-set (4 alt.)	25%	35%-65%
Luzhnica et al. (2016)	6	300 min	26 letters	98	Open-set	< 1%	(> 90%)*
Luzhnica & Veas (2019)	7	280 min	26 letters	98	Open-set	< 1%	(> 96%)*
Our Study II	24	100 min	39 phonemes	100	Closed-set	1%	80% - 92.5%
Our Study III (naïve Ps)	24	266-490 min	39 phonemes	500	Open-set	< 0.2%	65%

on the arms by Zhao *et al.* [41] and Turcott *et al.* [42], and one study that used radial squeeze and lateral skin stretch in addition to vibratory stimulation on the upper arm by Dunkelberger *et al.* [43]. These studies used 6 to 16 actuators, learning time varied from 30 to 100 minutes, and the number of phonemes encoded was 9 to 23. The two Facebook studies trained and tested the participants on 20 words. The Dunkelberger *et al.* study [43] trained on 150 words and tested with 50 words. During the word tests, Zhao *et al.* [41] required the participants to type the perceived word without looking at the word list. We call this open-set responses in that the participants were not limited in the words they selected for their responses. The corresponding chance level is listed as less than 1 divided by the number of word alternatives in the stimulus set, as the participants may not have memorized all the words. The other two studies used closed-set responses in that the participants chose their responses from a list of words; thus, the chance levels for word identification tests were simply 1 divided by the number of words in the test lists. It should also be mentioned that the Dunkelberger *et al.* study [43] allowed the participants to initiate the presentation of each phoneme in a word (i.e., resulting in a self-paced rate of phoneme presentation) instead of presenting the phonemes of a given word in a sequence with a pre-determined inter-phoneme interval. We discuss the issue of timing and presentation rate in another paper (see [44]). The three studies reported word recognition rates in the range of 76% to 87% correct that were significantly higher than chancel levels (2% to 5%). The results clearly indicate the feasibility of encoding speech on the skin. However, the number of phonemes employed in the three studies do not allow for the coding of all English words. There is the possibility that phoneme identification rate may deteriorate as more phonemes are introduced, possibly impacting word recognition accuracy. It is therefore unclear whether the same high level of word recognition could be maintained once the number of phonemes is increased to allow for any English word to be encoded and delivered with the haptic displays described in the three studies.

The remaining 5 studies listed in Table VI are based on approaches that can encode any English word. The list includes findings from three tactile speech communication systems: a tactile vest [45], the "skin reading" glove [46], [47], and our own TAPS (the studies presented in this article). Novich developed a haptic vest containing 27 tactors and tested a spectral-based approach with 50 words in a 4-alternative forced-choice identification paradigm (the correct answer was among the four response alternatives) [45]. After 11 to 12 days of training with 300 trials per day (training time was not reported), participants achieved scores of 35-65% correct (cf. chance performance of 25% correct on the task). It is unclear what the word identification score would have been if the participants in Novich's study had to choose one word from all 50 alternatives. Luzhnica *et al.*'s tactile glove used 5 tactors at the back of each digit and 1 tactor at the back of the palm to encode the 26 letters of the English alphabet [46]. After 300 min of training, the participants were tested with a stimulus set of 98 words and were instructed to type any English word as a response. Word recognition performance was calculated as the percentage of correctly entered letters rather than percentage of "whole" words. In a more recent study [47], Luzhnica & Veas added an additional tactor on the back of the palm and reported improved performance levels with an "optimized" set of vibrotactile codes for letters of the alphabet. This time word accuracy was computed by measuring the Levenshtein distance between the presented and recognized words [47]. The "skin reading" glove is relatively simple in terms of the number of tactors. The two studies show impressive performance of $>90\%$ accuracy calculated from letters, demonstrating the potential for the glove as a tactile speech communication device. One caveat is that the ability to "chunk" individual symbols into meaningful words is not a trivial process and in some instances may take months to achieve [48]. Therefore it remains to be seen whether the participants in Luzhnica *et al.*'s studies [46], [47] received meaningful word information, or if the same high performance level can be achieved when word accuracy is calculated on whole words.

In terms of our TAPS system, Study II and Study III both included all 39 phonemes and had the relatively large word vocabulary size of 100 and 500 words, respectively. Our participants responded by either selecting a word from a list (for the 100-word list) or typing any English word on a keyboard (for

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

the 500-word list). A response was counted as correct only if it matched or is a homonym of the word presented. In Study II, 8 of the 12 participants in the phoneme-based group achieved an average word recognition rate of 80% correct, and 2 of the 12 participants in the word-based group achieved an average of 92.5% correct, with a total learning time of 100 minutes. In Study III, 9 of the 10 naïve participants (excluding P46, Figure 9) achieved an average of 65% after 266-490 minutes of learning. Given the differences in vocabulary size and hence task difficulty, word-recognition percent-correct scores cannot be directly compared. In terms of the number of words learned (defined by word-recognition percent-correct score multiplied by number of words in the vocabulary), the results of Study II and Study III in this article represent the largest number of words that have been learned using a tactile speech communication device among the recent studies summarized in Table VI.

From a practical point of view, there are several factors that are necessary for any tactile speech device to become a useful means of communication. These include a reasonably short period of learning before the system can be used for meaningful interactions, as well as the need for a growth in vocabulary size with increased experience with the device. Throughout our project, we have been experimenting with the training curricula to gain insight on the best practices for helping users become proficient at receiving English words through TAPS. We found that the general principles of memory consolidation [34] and distributed practice [49] work well in the acquisition of a tactile language (see evidence in Fig. 5, [33]). Participants spent limited time per day and continued with the learning over several weeks, as opposed to devoting long hours in intensive training. In addition, we found it effective to interleave phoneme and word practices to gradually build up vocabulary size and competency [50]. In our experiments, we tracked learning time in minutes so that learning rates could be estimated. The data reported in this article show an average learning rate of roughly 1 English word per minute up to a vocabulary size of 500 words for the best performers. With a 500-word list, it is unlikely that the participants could memorize all the words. Our results therefore provide evidence of transmitting any English word through the skin with an open vocabulary.

VII. Conclusion and Future Work

We have developed a phonemic-based tactile speech communication system called TAPS for delivering speech to the skin. The feasibility of our phonemic-based approach was supported by the phoneme identification results in Reed *et al.* [30]. The present work focused on word recognition performance. Fifty-one participants took part in three studies with increasing number of phonemes and vocabulary sizes. Our results show that the best participants were able to learn up to 500 English words with a rate of roughly one English word per minute. The findings demonstrate the feasibility of transmitting (potentially) any English word using TAPS within a reasonable period of learning.

Ongoing and future work will proceed in several directions. To expand the capability of TAPS, we have implemented a text-to-speech front end to TAPS so written text can be automatically transcribed to phoneme streams. An automatic speech recognizer front end is also being implemented at this time so that any spoken English word can be readily presented via TAPS. This has enabled two highly-experienced participants to communicate with each other via text messages transmitted through two TAPS systems. Their performance will shed light on the information transmission rates that can be ultimately achieved through a tactile speech communication system. Another important research goal is concerned with the need for an increase in the speech transmission rate in an effort to match the 60-80 wpm rate demonstrated by Tadoma users. The results reported in this article are based on word transmission rates in the 30-40 wpm range. Increased communication rates may be accomplished by shortening the haptic symbols used to encode phonemes as well as by creating additional symbols to represent frequently-occurring phoneme pairs. There is also the need to embed learning in more engaging activities (such as games) to facilitate better learning outcomes. Future research will involve people with severe hearing impairments and assess their ability to use TAPS for speech communication. These activities will contribute towards a practical tactile speech communication system for people with all levels of sensory capabilities.

Acknowledgment

The authors would like to thank Freddy Abnousi, Frances Lau, Ali Israr, Keith Klumb, and Robert Turcott for their valuable insight, advice, and support of the studies presented here.

References

[1] C. M. Reed, N. I. Durlach, and L. D. Braida, “Research on tactile communication of speech: A review,” *Amer. Speech-Language-Hearing Assoc. (ASHA) Monographs Number 20*, vol. 23, 1982.
[2] C. M. Reed, N. I. Durlach, and L. A. Delhorne, “Natural methods of tactual communication,” in *Tactile Aids for the Hearing Impaired*, I. R. Summers (Ed.), London: Whurr Publishers, 1992, pp. 218–230.
[3] M. C. Schultz, S. J. Norton, S. Conway-Fithian, and C. M. Reed, “A survey of the use of the Tadoma method in the United States and Canada,” *Volta Rev.*, vol. 86, no. 6, pp. 282–292, 1984.
[4] S. J. Norton *et al.*, “Analytic study of the Tadoma method: Background and preliminary results,” *J. Speech Hearing Res.*, vol. 20, no. 3, pp. 574–595, 1977.
[5] C. M. Reed, N. I. Durlach, L. D. Braida, and M. C. Schultz, “Analytic study of the Tadoma method: Identification of consonants and vowels by an experienced Tadoma user,” *J. Speech Hearing Res.*, vol. 25, pp. 108–116, 1982.
[6] C. M. Reed, S. I. Rubin, L. D. Braida, and N. I. Durlach, “Analytic study of the Tadoma method: Discrimination ability of untrained observers,” *J. Speech Hearing Res.*, vol. 21, no. 4, pp. 625–637, 1978.
[7] C. M. Reed, M. J. Doherty, L. D. Braida, and N. I. Durlach, “Analytic study of the Tadoma method: Further experiments with inexperienced observers,” *J. Speech Hearing Res.*, vol. 25, pp. 216–223, 1982.
[8] C. M. Reed, W. M. Rabinowitz, N. I. Durlach, L. D. Braida, S. Conway-Fithian, and M. C. Schultz, “Research on the Tadoma method of speech communication,” *J. Acoustical Soc. Amer.*, vol. 77, no. 1, pp. 247–257, 1985.
[9] W. M. Rabinowitz, D. R. Henderson, C. M. Reed, L. A. Delhorne, and N. I. Durlach, “Continuing evaluation of a synthetic Tadoma system,” *J. Acoustical Soc. Amer.*, vol. 87, no. S1, pp. S88–S88, 1990.
[10] A. Carrera, A. Alonso, R. D. l. Rosa, and E. J. Abril, “Sensing performance of a vibrotactile glove for deaf-blind people,” *Appl. Sciences*, vol. 7, no. 4, 2017, Art. no. 317.

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

[11] C. M. Reed, "Tadoma: An overview of research (chapter 4)," in *Profound Deafness and Speech Communication*, G. Plant and K.-E. Spens, Eds., London: Whurr Publishers, 1995, pp. 40–55.

[12] J. C. Krause and L. D. Braida, "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," *J. Acoustical Soc. Amer.*, vol. 112, no. 5, pp. 2165–2172, 2002.

[13] C. M. Reed, N. I. Durlach, L. A. Delhorne, W. M. Rabinowitz, and K. W. Grant, "Research on tactual communication of speech: Ideas, issues, and findings," *Volta Rev.*, vol. 91, pp. 65–78, 1989.

[14] L. Duchesne, A. Sutton, and F. Bergeron, "Language achievement in children who received cochlear implants between 1 and 2 years of age: Group trends and individual patterns," *J. Deaf Stud. Deaf Educ.*, vol. 14, no. 4, pp. 465–85, 2009.

[15] M. A. Svirsky, A. M. Robbins, K. I. Kirk, D. B. Pisoni, and R. T. Miyamoto, "Language development in profoundly deaf children with cochlear implants," *Psychological Sci.*, vol. 11, no. 2, pp. 153–158, 2000.

[16] R. S. Cowan, J. I. Alcantara, L. A. Whitford, P. J. Blamey, and G. M. Clark, "Speech perception studies using a multichannel electrotactile speech processor, residual hearing, and lipreading," *J. Acoustical Soc. Amer.*, vol. 85, no. 6, pp. 2593–2607, 1989.

[17] I. R. Summers, "Information transfer through the skin: Limitations and possibilities," *Les Cahiers De l'Audition*, vol. 13, pp. 34–37, 2000.

[18] K. E. Spens and G. Plant, "A tactual "hearing" aid for the deaf," *STL-QPSR*, vol. 24, no. 1, pp. 52–56, 1983.

[19] D. Franklin, "Tactile aids, new help for the profoundly deaf," *Hearing J.*, vol. 37, no. 2, pp. 20–23, 1984.

[20] C. M. Reed and L. A. Delhorne, "Current results of field study of adult users of tactile aids," *Seminars Hearing*, vol. 16, no. 4, pp. 305–315, 1995.

[21] J. M. Weisenberger and M. E. Percy, "The transmission of phoneme-level information by multichannel tactile speech perception aids," *Ear Hearing*, vol. 16, no. 4, pp. 392–406, 1995.

[22] M. F. Nolan, "Two-point discrimination assessment in the upper limb in young adult men and women," *Physical Therapy*, vol. 62, no. 7, pp. 965–969, 1982.

[23] R. W. Cholewiak and A. A. Collins, "Vibrotactile localization on the arm: Effects of place, space, and age," *Perception Psychophysics*, vol. 65, no. 7, pp. 1058–1077, 2003.

[24] D. Wang, C. Peng, N. Afzal, W. Li, D. Wu, and Y. Zhang, "Localization performance of multiple vibrotactile cues on both arms," *IEEE Trans. Haptics*, vol. 11, no. 1, pp. 97–106, Jan.-Mar. 2018.

[25] F. A. Geldard, "Cutaneous coding of optical signals: The Optohapt," *Perception Psychophysics*, vol. 1, pp. 377–381, 1966.

[26] A. Gallace, H. Z. Tan, and C. Spence, "Numerosity judgments for tactile stimuli distributed over the body surface," *Perception*, vol. 35, no. 2, pp. 247–266, 2006.

[27] "The Braille Literacy Crisis in America: Facing the Truth, Reversing the Trend, Empowering the Blind," National Federation of the Blind, 2009.

[28] H. Z. Tan, N. I. Durlach, W. M. Rabinowitz, C. M. Reed, and J. R. Santos, "Reception of Morse code through motional, vibrotactile, and auditory stimulation," *Perception Psychophysics*, vol. 59, no. 7, pp. 1004–1017, 1997.

[29] W. D. Keidel, "The cochlear model in skin stimulation," in *Cutaneous Communication Systems and Devices*, F. A. Geldard, Ed., Monterey, CA: The Psychonomic Society, Inc., 1973, pp. 27–32.

[30] C. M. Reed *et al.*, "A phonemic-based tactile display for speech communication," *IEEE Trans. Haptics*, vol. 12, no. 1, pp. 2–17, Jan.-Mar. 2019.

[31] D. H. Ecroyd, M. M. Halfond, and C. C. Towne, Eds. *Voice and Articulation: A Handbook*. Glenview, IL: Scott, Foresman and Co., 1966.

[32] L. A. Jones and H. Z. Tan, "Application of psychophysical techniques to haptic research," *IEEE Trans. Haptics*, vol. 6, no. 3, pp. 268–284, Jul.-Sept. 2013.

[33] J. Jung *et al.*, "Speech communication through the skin: Design of learning protocols and initial findings," in *Proc. Int. Conf. Des., User Experience, Usability*, pp. 447–460, 2018.

[34] Y. Dudai, A. Karni, and J. Born, "The consolidation and transformation of memory," *Neuron*, vol. 88, no. 1, pp. 20–32, 2015.

[35] Y. Jiao *et al.*, "A comparative study of phoneme- and word-based learning of English words presented to the skin," *Proc. EuroHaptics 2018*, vol. LNCS 10894, pp. 623–635, 2018.

[36] S. Engelmann and R. Rosov, "Tactual hearing experiment with deaf and hearing subjects," *J. Exceptional Children*, vol. 41, no. 4, pp. 243–253, 1975.

[37] P. L. Brooks and B. J. Frost, "Evaluation of a tactile vocoder for word recognition," *J. Acoustical Soc. Amer.*, vol. 74, no. 1, pp. 34–39, 1983.

[38] P. L. Brooks, B. J. Frost, J. L. Mason, and K. Chung, "Acquisition of a 250-word vocabulary through a tactile vocoder," *J. Acoustical Soc. Amer.*, vol. 77, no. 4, pp. 1576–1579, 1985.

[39] M. P. Lynch, R. E. Eilers, D. K. Oller, and L. Lavoie, "Speech perception by congenitally deaf subjects using an electrocutaneous vocoder," *J. Rehabil. Res. Develop.*, vol. 25, no. 3, pp. 41–50, 1988.

[40] K. L. Galvin, P. J. Blamey, M. Oerlemans, R. S. Cowan, and G. M. Clark, "Acquisition of a tactile-alone vocabulary by normally hearing users of the Tickle Talker^{TM}," *J. Acoustical Soc. Amer.*, vol. 106, no. 2, pp. 1084–1089, 1999.

[41] S. Zhao, A. Israr, F. Lau, and F. Abnousi, "Coding tactile symbols for phonemic communication," in *Proc. 2018 ACM CHI Conf. Human Factors Comput. Syst.*, vol. 392, pp. 1–13, 2018.

[42] R. Turcott *et al.*, "Efficient evaluation of coding strategies for transcutaneous language communication," in *Proc. EuroHaptics 2018 (Springer LNCS 10894)*, pp. 600–611, 2018.

[43] N. Dunkelberger *et al.*, "Conveying language through haptics: A multi-sensory approach," in *Proc. 2018 ACM Int. Symp. Wearable Computers*, pp. 25–32, 2018.

[44] C. M. Reed, H. Z. Tan, Y. Jiao, Z. D. Perez, and E. C. Wilson, "Identification of words and phrases through a phonemic-based haptic display: Effects of inter-phoneme and inter-word interval durations," submitted to *ACM Transactions on Applied Perception, in review*.

[45] S. Novich, "Sound-to-Touch Sensory Substitution and Beyond," PhD dissertation, Dept. Elect. Comput. Eng. Rice Univ., Houston, Texas, 2015.

[46] G. Luzhnica, E. Veas, and V. Pammer, "Skin reading: Encoding text in a 6-channel haptic display," in *Proc. Int. Symp. Wearable Computers (ISWC) 2016*, pp. 148–155, 2016.

[47] G. Luzhnica and E. Veas, "Optimising encoding for vibrotactile skin reading," in *Proc. 2019 CHI Conf. Human Factors Comput. Syst.*, 2019, Art. no. 235.

[48] W. L. Bryan and N. Harter, "Studies on the telegraphic language: The acquisition of a hierarchy of habits," *Psychological Rev.*, vol. 6, no. 4, pp. 345–375, 1899.

[49] N. L. Foster, M. L. Mueller, C. Was, K. A. Rawson, and J. Dunlosky, "Why does interleaving improve math learning? The contributions of discriminative contrast and distributed practice," *Memory Cognition*, vol. 47, no. 6, pp. 1088–1101, 2019.

[50] D. Rohrer, R. F. Dedrick, and S. Stershic, "Interleaved practice improves mathematics learning," *J. Educational Psychol.*, vol. 21, pp. 1323–1330, 2015.

Hong Z. Tan (Fellow, IEEE) received the bachelor's degree in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1986 and the master's and doctorate degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1988 and 1996, respectively. She is currently a Professor of Electrical and Computer Engineering, Mechanical Engineering (by courtesy) and Psychological Sciences (by courtesy) with Purdue University, West Lafayette, IN, USA. She has served three terms as an Associate Editor for the IEEE TRANSACTIONS ON HAPTICS (2007 to 2012; 2016 to 2019) and received a Meritorious Service Award, in 2012.

Charlotte M. Reed received the bachelor's of science degree in education from Carlow College, Carlow, Ireland, in 1969, and the Ph.D. degree in bioacoustics from the University of Pittsburgh, Pittsburgh, PA, USA, in 1973. She is currently a Senior Research Scientist in the Research Laboratory of Electronics with the Massachusetts Institute of Technology, Cambridge, MA.

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.

Yang Jiao received the B.E. degree in electronic information science and technology from the Beijing University of Posts and Telecommunications, Beijing, China, in 2009, the M.S. degree in wireless communications from the University of Southampton, Southampton, U.K., in 2010, and the Ph.D. degree in design from Tsinghua University, Beijing, China, in 2017. He was a Postdoctoral Research Associate with Purdue University, West Lafayette, IN, USA.

Zachary D. Perez received the bachelor's degree in chemistry from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2005 and the J.D. degree from Georgetown University Law Center, Washington, DC, USA, in 2012. He was a Senior Research Support Associate in the Research Laboratory of Electronics with the Massachusetts Institute of Technology, Cambridge, MA, USA.

E. Courtenay Wilson received the bachelor's of science degree in computer science and engineering from the University of Connecticut, Storrs, CT, USA, in 1994, the master's of science degree in computer engineering from the University of Nevada, Reno, NV, USA, in 2001, and the Ph.D. degree in speech and hearing bioscience and technology from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2010. She is currently a Researcher with MIT and a Lecturer with the Math department, Northeastern University, Boston, MA, USA.

Jaehong Jung received the B.S. degree and is working toward the M.S. degree with the School of Mechanical Engineering, Purdue University, West Lafayette, IN, USA. He is currently a Research Assistant with the Haptic Interface Research Laboratory, Purdue University, under the supervision of Dr. Hong Z. Tan.

Juan S. Martinez received the B.A.Sc. degree in electronic engineering in 2016, the B.A.Sc. degree in systems and computer engineering, in 2017, from fromLos Andes University, Bogota, Colombia, and the M.S. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2019, where he is currently working toward the Ph.D. degree. He currently works as a Research Assistant with the Haptic Interface Research Laboratory, advised by Dr. Hong Z. Tan.

Frederico M. Severgnini received the B.A.Sc. degree in electronic and telecommunication engineering from PUC Minas University, Belo Horizonte, Brazil, in 2015, and the M.S. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2018. He worked as a Graduate Research Assistant with the Haptic Interface Research Laboratory, advised by Dr. Hong Z. Tan.

Authorized licensed use limited to: Purdue University. Downloaded on December 26,2020 at 15:19:47 UTC from IEEE Xplore. Restrictions apply.